J. Mol. Biol. (1985) **185**, 219–226

Refinement of the Solution Structure of the *B* DNA Hexamer 5′d(C-G-T-A-C-G)$_2$ on the Basis of Inter-proton Distance Data

A restrained least-squares refinement of the solution structure of the self-complementary *B* DNA hexamer 5′d(C-G-T-A-C-G)$_2$ is presented. The structure is refined on the basis of 190 inter-proton distances determined by pre-steady-state nuclear Overhauser enhancement measurements. Two refinements were carried out starting from two initial *B* DNA structures differing by an overall root-mean-square (r.m.s.) difference of 0·32 Å. In both cases, the final r.m.s. difference between the experimental and calculated inter-proton distances was 0·12 Å compared to 0·61 Å and 0·58 Å for the two initial structures. The difference between the two refined structures is small, with an overall r.m.s. difference of 0·16 Å, and represents the error in the refined co-ordinates. The refined structures have a *B*-type conformation with local structural variations in backbone and glycosidic bond torsion angles, and base-pair propellor twist, base roll, base tilt and local helical twist angles.

The crystal structure of the *B* DNA dodecamer solved by Dickerson & Drew (1981) revealed that *B* DNA is not a regular helix but rather exhibits a large degree of local structural variation. Given that crystal structures of oligonucleotides are subject to strong intermolecular interactions, in particular crystal packing forces and local high ionic conditions (Rhodes, 1982), it is of interest to develop approaches towards determining the three-dimensional structures of oligonucleotides in solution. Recently, we carried out a detailed nuclear magnetic resonance study on the self-complementary DNA hexamer 5′d(C-G-T-A-C-G)$_2$ under conditions in which this oligonucleotide is entirely double-stranded and determined a large number of intra- and internucleotide inter-proton distances using pre-steady-state nuclear Overhauser enhancement measurements (Gronenborn *et al.*, 1984). The inter-proton distance data were found to be indicative of a right-handed *B*-type structure with a mononucleotide repeating unit, in agreement with the *B*-type circular dichroism spectrum of the hexamer. In this letter we present the refinement of the solution structure of the *B* DNA hexamer on the basis of 190 inter-proton distances using a restrained least-squares refinement procedure.

The least-squares refinement program used was RESTRAIN (Haneef *et al.*, 1983; Haneef *et al.*, 1985). The function minimized in Cartesian co-ordinate space is given by:

$$C = \Sigma W_d(d_t - d_c)^2 + \Sigma W_v|V| + \Sigma W_b(b_o - b_{min})^2 \quad (b_o < b_{min}), \quad (1)$$

where W_d, W_v and W_b are weighting coefficients, d_t and d_c are the target and calculated inter-atomic distances, respectively, $|V|$ is the determinant of the product-moment matrix of planar groups of atoms, and b_o and b_{min} are the observed and minimum allowed distances between two non-bonded atoms. The inter-atomic distances include all distances between covalently bonded atoms, between atoms defining fixed bond angles, and between atoms defining hydrogen bonding in the A·T and G·C base-pairs, as well as the inter-proton distances determined from the pre-steady NOE† measurements. For each residue, the C-1′ atom of the deoxyribose and all the atoms of the base (with the exception of the methyl protons) are constrained to lie in the same plane. The last term in equation (1) is simply used to prevent undesirably close contacts and only comes into operation when $b_o < b_{min}$; this repulsive term was used only in the initial stages of the refinement.

The pre-steady-state NOE reasurements made by Gronenborn *et al.* (1984) were carried out under conditions in which the errors in the estimation of the inter-proton distances (with the exception of the internucleotide $r_{H8/H6-H8/H6}$ distances) are $\lesssim \pm 0{\cdot}2$ Å (Clore & Gronenborn, 1984, 1985*a*). The internucleotide $r_{H8/H6-H8/H6}$ distances, however, are subject to severe underestimation (by $\lesssim 0{\cdot}5$ Å) because the NOE values between these protons are very small ($\sim -2\%$) and contain a significant contribution from indirect cross-relaxation *via* the H2′/H2″ sugar protons, on account of their close proximity to both the H8/H6 proton of their own residue and the H8/H6 proton of the adjacent 3′ residue. For this reason, the internucleotide $r_{H8/H6-H8/H6}$ distances were not included in the refinement presented here.

A total of 95 of the experimentally determined inter-proton distances yielded 190 distance

† Abbreviations used: NOE, nuclear Overhauser effect; r.m.s., root-mean-square.

0022-2836/85/170219-08 $03.00/0

restraints in the self-complementary hexamer and covered the range 2·1 to 3·8 Å. An additional 32 restraints (see Table 1) define the A·T and G·C base-pairing giving a total of 222 restraints in the hexamer. These were used in the least-squares refinement of the Cartesian co-ordinates of the atoms, together with further restraints that were used to preserve the known covalent geometry of the molecule. The 222 restraints essentially determine the 78 torsion angles (the glycosidic bond (χ) and the main-chain torsion angles (α to ξ)) that define the conformation of the two strands of the molecule.

Two refinements were carried out. For the first refinement, the starting co-ordinates, known as initial *B* DNA model I, were those of classical *B* DNA derived from the fibre diffraction data of Arnott & Hukins (1972). For the second refinement, the starting co-ordinates, known as initial *B* DNA model II, were obtained by subjecting the co-ordinates for initial *B* DNA model I to 500 cycles of energy minimization using the program CHARMM (Brooks *et al.*, 1983). This structure is very similar to the energy-refined *B* DNA structure obtained by Levitt (1978). These two starting structures are entirely reasonable as both circular dichroism (Kuzmich *et al.*, 1982) and NOE (Gronenborn *et al.*, 1984) data have shown that the structure of the hexamer in solution is that of right-handed *B* DNA.

In order that the distance restraints should be weighted so as to reflect their estimated precision, advantage was taken of a facility in RESTRAIN, which is used in the X-ray refinement of macromolecules and allows distances to be classified into three ranges: $r < 2{\cdot}12$ Å, $2{\cdot}12$ Å $< r < 2{\cdot}62$ Å and $r > 2{\cdot}62$ Å. The weights applied in these ranges were in the ratio 5 : 4 : 3. Application of restraints to non-bonded contacts was undertaken only in the first five cycles of refinement. A total of 30 cycles of refinement were performed. The distance weights were chosen to represent approximately the gradation of error as a function of distance for the experimental inter-proton distances (Clore & Gronenborn, 1985*a*). In this respect, it is important to bear in mind that although the terms in equation (1) can be considered as pseudo-potentials, the refinement carried out here is not an energy refinement. Thus, the change in conformation on refinement arises solely from the inter-proton distances restraints, as all the other restraints are

Table 1

r.m.s. difference (Å) between the target restraints and the corresponding calculated values in the initial B *DNA models and the final refined structures of* $5'd(C\text{-}G\text{-}T\text{-}A\text{-}C\text{-}G)_2$

	Number of restraints	r.m.s. difference (Å) Initial *B* DNA models[a] I	II	Final refined structures I	II
All distance restraints					
$r < 2{\cdot}12$ Å	538	0·028	0·027	0·029	0·032
$2{\cdot}12$ Å $r < 2{\cdot}62$ Å	690	0·156	0·153	0·076	0·078
$r > 2{\cdot}62$ Å	112	0·680	0·652	0·112	0·111
Covalent and bond angle restraints[b]					
$r < 2{\cdot}12$ Å	522	0·017	0·018	0·029	0·033
$2{\cdot}12$ Å $< r < 2{\cdot}62$ Å	584	0·033	0·038	0·065	0·067
$r > 2{\cdot}62$ Å	12	0·001	0·040	0·038	0·041
Planes[c]	12	0·003	0·017	0·022	0·017
Base-pairing restraints[d]	32	0·000	0·106	0·033	0·031
Inter-proton distances[e]	190	0·605	0·576	0·124	0·123
Total number of atoms: 374 (1116 degress of freedom)					
Total number of restraints: 1352					

[a]Initial *B* DNA model I is derived from the fibre diffraction data of Arnott & Hukins (1972). Initial *B* DNA model II was obtained by subjecting initial *B* DNA model I to 500 cycles of energy minimization using the program CHARMM (Brooks *et al.*, 1983).

[b]It should be noted that bond angles are defined by inter-atomic distances.

[c]For each residue the C-1′ atom of the deoxyribose and all the atoms of the base (with the exception of the methyl protons) are constrained to lie in the same plane.

[d]The base-pairing restraints are as follows: for A·T base-pairs, $r_{A(N6)-T(O4)} = 2{\cdot}78$ Å, $r_{A(HN')-T(O4)} = 1{\cdot}70$ Å, $r_{A(N1)-T(H3)} = 1{\cdot}74$ Å and $r_{A(N1)-T(N3)} = 2{\cdot}82$ Å; for G·C base-pairs, $r_{G(H1)-C(HN')} = 1{\cdot}62$ Å, $r_{G(O6)-C(N4)} = 2{\cdot}70$ Å, $r_{G(N1)-C(N3)} = 2{\cdot}82$ Å, $r_{G(H1)-C(N3)} = 2{\cdot}82$ Å, $r_{G(H1)-C(N3)} = 1{\cdot}74$ Å, $r_{G(N2)-C(O2)} = 2{\cdot}78$ Å and $r_{G(HN')-C(O2)} = 1{\cdot}70$ Å. These values are those of initial *B* DNA model I derived from the fibre diffraction data of Arnott & Hukins (1972).

[e]These inter-proton distances are those determined by Gronenborn *et al.* (1984) using pre-steady-state NOE measurements. They do not include inter-proton distances that are fixed by the geometry of the sugar ring and bases themselves.

Table 2

Overall r.m.s. shifts in Cartesian co-ordinates and r.m.s. shift in glycosidic and main-chain torsion angles between the initial and final refined structures of 5'd(C-G-T-A-C-G)$_2$

	Overall r.m.s. shift (Å)	r.m.s. difference in glycosidic and main-chain torsion angles (°)
Initial I *versus* initial II	0·32	15·4
Initial I *versus* refined I	0·54	27·3
Initial I *versus* refined II	0·60	30·0
Initial II *versus* refined II	0·55	19·1
Initial II *versus* refined I	0·56	17·2
Refined I *versus* refined II	0·16	4·2

well satisfied in the initial and final structures as well as in the intermediate structures sampled during the entire course of the refinement.

The r.m.s. difference between the target and calculated values for the distance and planarity restraints in the initial and final structures is given in Table 1; the overall r.m.s. shifts in Cartesian co-ordinates and the r.m.s. differences in glycosidic and main-chain torsion angles between the structures are given in Table 2; and the conformational parameters describing the initial and refined structures are given in Tables 3 and 4. The average r.m.s. differences in the co-ordinates of the sugar–phosphate and base moieties between the structures are plotted as a function of residue number in Figure 1; stereo views of refined structure I in the form of skeletal and space-filling diagrams are shown in Figure 2(a) and (b), respectively; Figure 2(c) shows a stereo view of the initial *B* DNA model I and the refined structure I superimposed; a stereo view of refined structures I and II superimposed is shown in Figure 3; and stereo views of the five individual base-pair steps of refined structure I viewed down the helix axis are shown in Figure 4.

It is clear from the data in Table 1 that the refinement has resulted in a considerable improvement in the agreement between calculated and target inter-proton distances with an r.m.s. difference of 0·12 Å for the final refined structures, which is within the error of the data (±0·2 Å), compared to values of 0·61 Å and 0·58 Å for initial *B* DNA models I and II, respectively. At the same time, all the other restraints are well satisfied in the refined structures and there are no undesirably close non-bonded contacts. It can also be seen from the data in Table 2 and Figures 1 and 3 that the

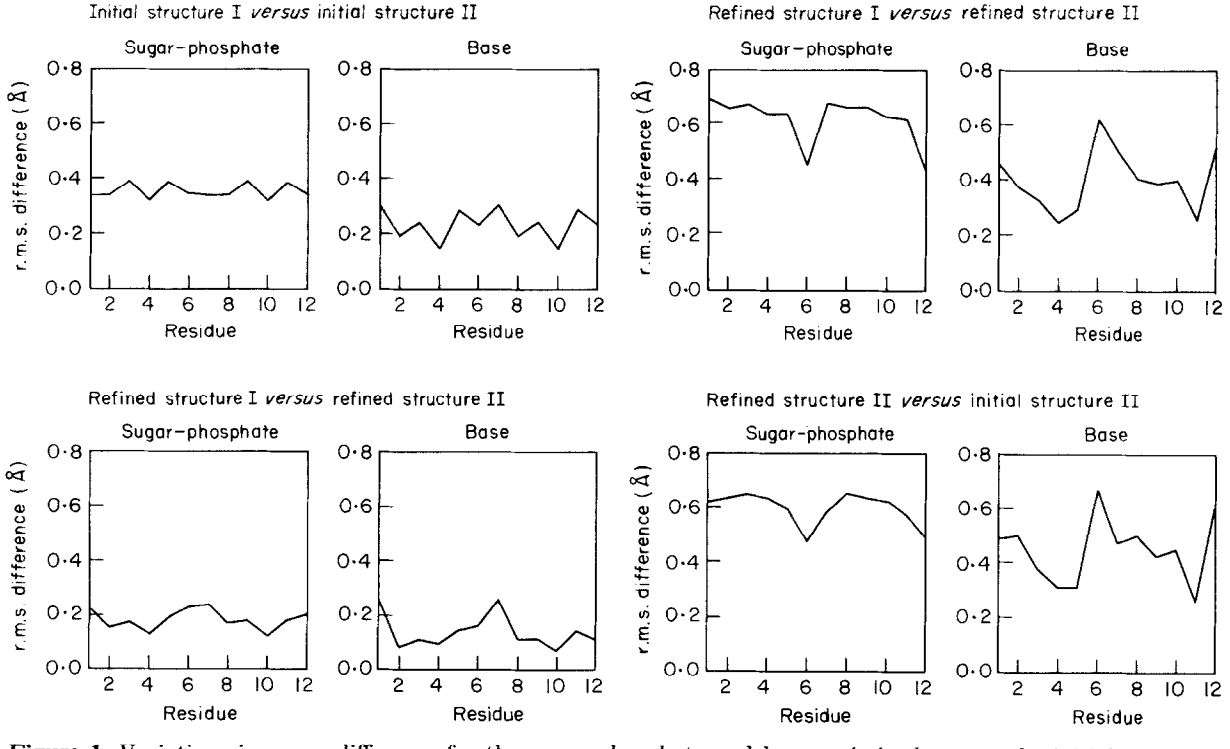

Figure 1. Variations in r.m.s. difference for the sugar–phosphate and base moieties between the initial and refined structures of 5'd(C-G-T-A-C-G)$_2$ as a function of residue number.

Table 3

Torsion and propellor twist angles for refined structures I and II of 5′d(C-G-T-A-C-G)$_2$

Residue	Glycosyl(°) χ	Refined structure I/refined structure II Main-chain torsion angles (°) α	β	γ	δ	ε	ζ	Adjacent phosphorus atom separation (Å)	Observed sugar conformation	Base-pair propellor twist ψ(°)
C_1, C_7	−114/−128	—	−154/−161	67/67	114/108	153/155	−69/−69	6·42/6·55	C-1′-exo	6/6
G_2, G_8	−111/−116	−77/−81	−146/−153	63/63	112/110	142/139	−65/−63	6·34/6·45	C-1′-exo	14/12
T_3, T_9	−117/−121	−86/−93	−146/−152	66/68	112/108	138/138	−62/−62	6·32/6·43	C-1′-exo	6/8
A_4, A_{10}	−117/−118	−90/−94	−140/−143	65/67	113/110	143/142	−67/−65	6·54/6·62	C-1′-exo	6/8
C_5, C_{11}	−113/−116	−82/−90	−146/−150	61/64	115/109	136/138	−62/−64	6·35/6·46	C-1′-exo	11/13
G_6, G_{12}	−96/−99	−87/−91	−136/−141	65/65	124/124	—	—	—	C-1′-exo	6/6
Mean	−114±9	−87±6	−147±7	65±2	113±5	142±7	−65±3	6·45±0·1	C-1′-exo	8·5±3
B DNA I[a]	−98	−47	−146	36	156	155	−95	6·46	C-2′-endo	0
B DNA II[a]	−107±4	−66±1	−157±2	58±2	133±3	163±1	−91±1	6·62±0·1	C-2′-endo	5±2
B_C DNA[b]	−117±14	−63±8	171±14	54±8	123±21	−169±25	−108±34	6·68±0·23	O-1′-endo to C-2′-endo with mean C-1′-exo	13±5
B_S DNA[c]	−113±14	−61±16	−153±11	38±14	129±17	155±13	−71±23	6·42±0·18	C-1′-exo	10±7

The main-chain torsion angles are defined by P^{α}-O-5′$^{\beta}$-C-5′$^{\gamma}$-C-4′$^{\delta}$-C-3′$^{\varepsilon}$-O-3′$^{\zeta}$P and the glycosidic bond torsion angles are defined by χ_{pur} = O-1′–C-1′–N-9–C-4 and χ_{pyr} = O-1′–C-1′–N–1–C-2 with zero at the eclipsed position and positive angles by clockwise rotation of the further pair of atoms. The propellor twist angle ψ is the dihedral angle between individual base planes. The values of the conformational parameters given are the means of the values for the 2 strands of the hexamer. The r.m.s. difference between the values of the glycosidic and main-chain torsion angles for strands 1 and 2 is 1·1° and 1·4° for refined structures I and II, respectively.

[a]Parameters for initial *B* DNA models I and II.

[b]From the crystal data of Dickerson & Drew (1981) on the self-complementary dodecamer 5′d(C-G-C-G-A-A-T-T-C-G-C-G)$_2$.

[c]From the data of Clore & Gronenborn (1985*b*) on the refined solution structure of the undecamer [5′d(A-A-G-T-G-T-G-A-C-A-T) · 5′d(A-T-G-T-C-A-C-A-C-T-T)].

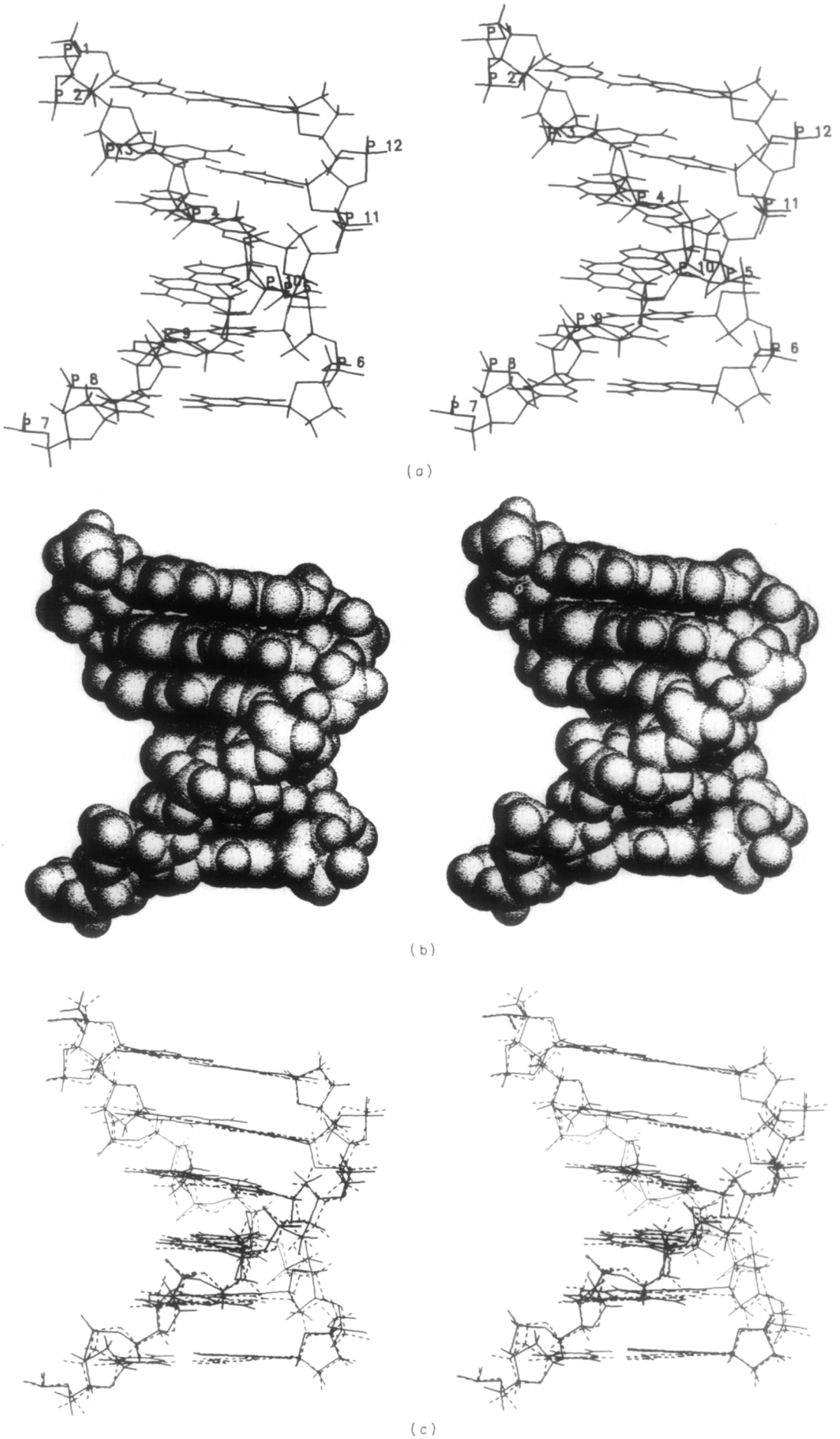

Figure 2. Skeletal (a) and space-filling (b) stereo views of refined structure I of 5′d(C-G-T-A-C-G)$_2$. (c) Stereo view of the initial classical *B* DNA model I (broken lines) and the final refined structure I (continuous lines).

Table 4

Local helical twist (t_1), base roll (θ_R) and base tilt (θ_T) angles for the refined structures I and II of 5'd(C-G-T-A-C-G)$_2$

Base-pair step	Refined structure I/refined structure II t_1(°)	θ_R(°)	θ_T(°)
1 C·G	34/36	−3/−3	1/5
2 G·C	33/33	0/−1	2/3
3 T·A	36/38	−2/−1	1/4
4 A·C	33/32	−7/−6	1/2
5 C·G	35/36	0/−1	3/4
Mean	34·6±1·9	−2·4±2·4	2·6±1·4
B DNA I[a]	36		
B DNA II[a]	35·8±0·4	−2·1±4·4	−0·1±2·7
B_C DNA[b]	37·3±3·8	0·24±6·7	4·0±3·8
B_S DNA[c]	35±3	−2·3±8·8	1·4±5·2

The base roll angle θ_R is the rotation about an axis in the plane of the bases perpendicular to the pseudo-dyad and is positive when opening towards the minor groove.

The base tilt angle θ_T is the rotation about the pseudo-dyad axis passing through the base plane and is positive when opening to the outside of the molecule.

[a]Parameters for initial *B* DNA models I and II.

[b]From the crystal data of Dickerson & Drew (1981).

[c]For the refined solution structure of a *B* DNA undecamer (Clore & Gronenborn, 1985*b*).

difference between the two refined structures is insignificant (0·16 Å) and, in terms of r.m.s. difference in Cartesian co-ordinates, is a factor of 2 less than the difference between the two starting structures (0·32 Å). The very close similarity between the two refined structures and the much larger difference between the two starting structures are emphasized further by a comparison of the difference in glycosidic bond and backbone torsion angles (see Tables 2 and 3). Given that initial model II was derived from initial model I by energy minimization, it is also important to assess whether initial model II represents an intermediate structure along the first refinement pathway from initial model I to refined structure I. That this is *not* the case is easily ascertained by two different approaches. First, we note that the r.m.s. differences in Cartesian co-ordinates between the initial and final structures for all four possible combinations are approximately the same, ranging from 0·52 to 0·60 Å. Given that the r.m.s. difference between the intermediate structures and the final refined structure is reduced at every successive step of the refinement, it follows that the two refinements starting from initial models I and II represent independent pathways and that initial model II cannot be sampled along the first refinement pathway. This is confirmed by an analysis of the r.m.s. difference in glycosidic bond and main-chain torsion angles between initial model II and the intermediate structures generated during the course of the first refinement. At no point does this torsion angle r.m.s. difference fall below 12° (the minimum being reached at cycle 7) from a starting value of 15° to a final value of 17°.

The solution structure of the hexamer is clearly not rigid and static but dynamic, and indeed cross-relaxation measurements between protons a fixed distance apart have demonstrated the presence of internal mobility, particularly in the deoxyribose moieties (Clore & Gronenborn, 1984). Consequently, the refined structures should be viewed as representations of an "average" structure about which fluctuations can take place. In this respect it is important to remember that the experimentally measured inter-proton distances are not arithmetic means but $(\langle r^{-6} \rangle)^{-1/6}$ means so that they are heavily weighted towards fluctuations with the shortest inter-proton distances. This could potentially result in a distorted refined structure in the presence of large-magnitude internal motions. However, many of the inter-proton distances are highly correlated so that a single structure would

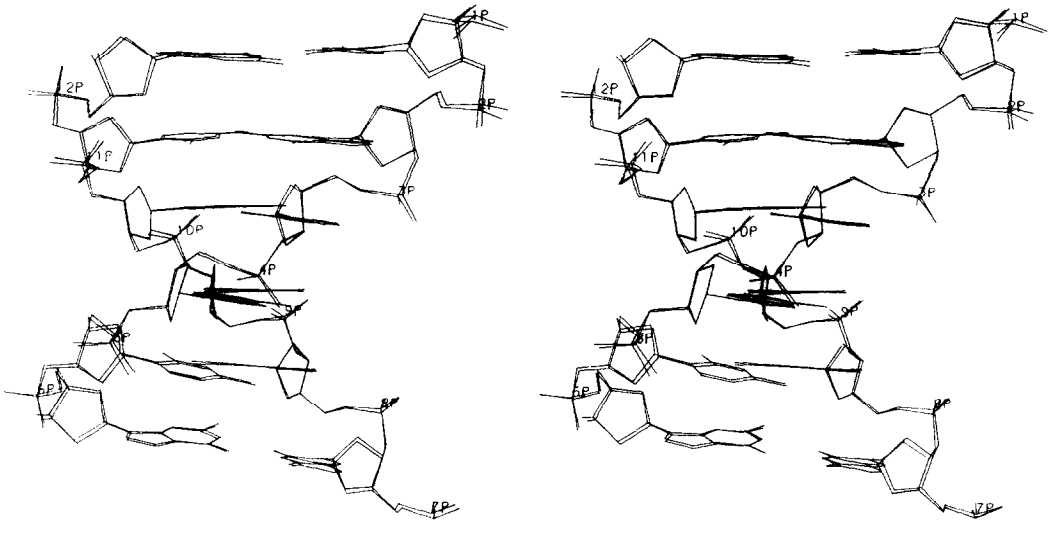

Figure 3. Stereo view of refined structures I and II of 5'd(C-G-T-A-C-G)$_2$ superimposed. For the sake of clarity protons have been omitted.

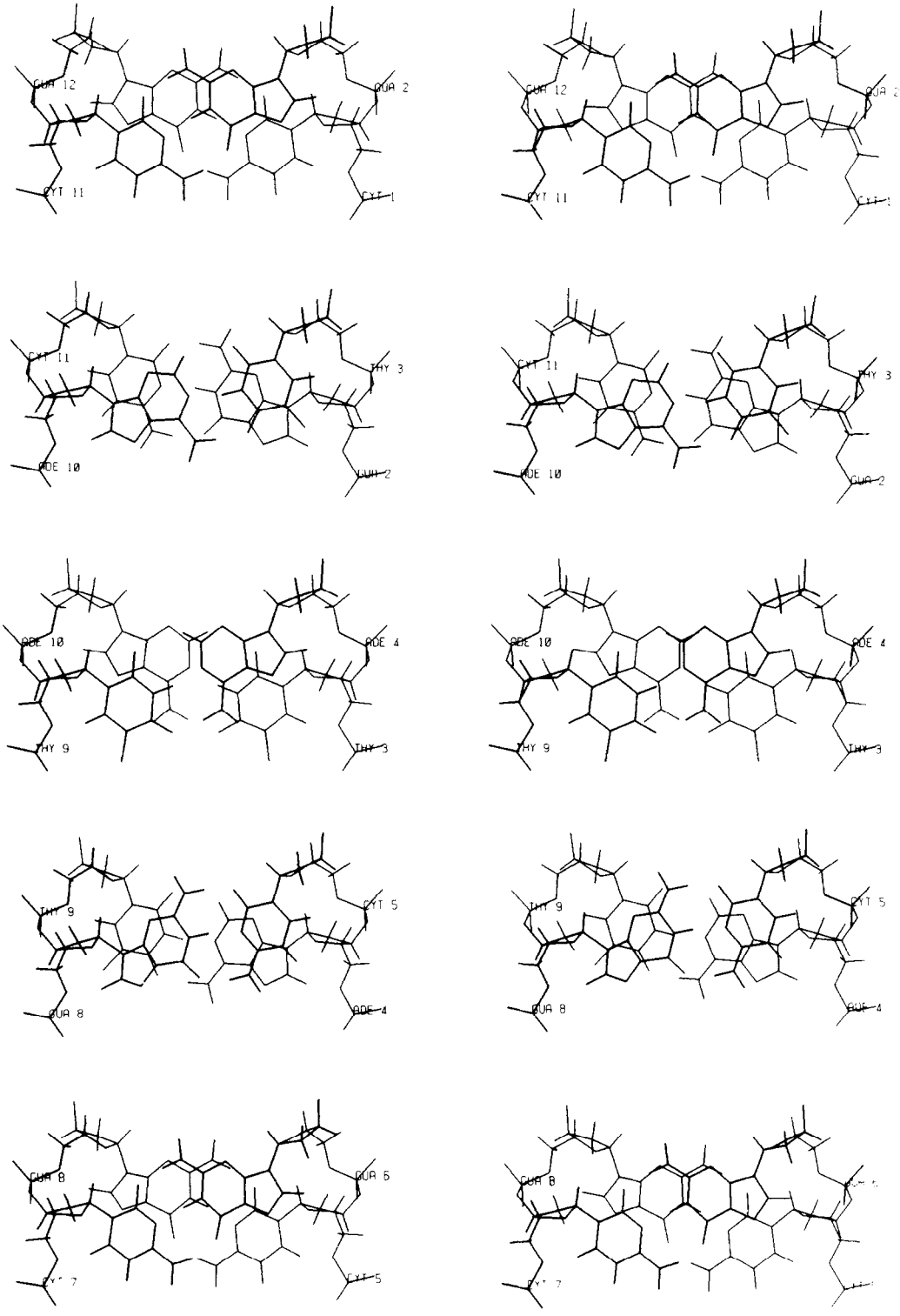

Figure 4. Stereo drawings of the 5 individual base-pair steps of refined structure I of 5′d(C-G-T-A-C-G)$_2$ viewed down the helix axis.

not be able to provide an acceptable fit to the experimental data if the magnitude of the internal motions were large (Clore & Gronenborn, 1985*b*). In r.m.s. difference of only 0·12 Å between the experimental and calculated inter-proton distances for the two refined structures. We therefore conclude that refined structures I and II are good representations of the "true" solution structure and refined co-ordinates.

Looking at Figures 2 and 3 as well as the

conformational parameters in Tables 3 and 4, it is clear that the overall *B*-type conformation is preserved in the refined structure of the self-complementary hexamer with the main-chain torsion angles α to ξ exhibiting the conformations g^-, t, g^+, t, t, g^- as expected. Like the crystal structure of the self-complementary *B* DNA dodecamer 5′d(C-G-C-G-A-A-T-T-C-G-C-G)$_2$ (Dickerson & Drew, 1981) and the refined solution structure of the *B* DNA undecamer [5′d(A-A-G-T-G-T-G-A-C-A-T) · 5′d(A-T-G-T-C-A-C-A-C-T-T)] (Clore & Gronenborn, 1985*b*), however, the refined structure of the hexamer is no longer a regular helix but exhibits local structural variations.

There are several features of the refined structures of the hexamer that deserve mentioning.

(1) The refined structures are not perfectly symmetric. This departure from perfect symmetry is not surprising as, once asymmetry is introduced during the refinement, for example as a result of rounding errors, it will persist. Nevertheless, the deviations from symmetry are very small (with an r.m.s. difference between the glycosidic and main-chain torsion angles for the two strands of 1·1° and 1·4° for refined structures I and II, respectively), in marked contrast to those observed in the crystal structure of the self-complementary *B* DNA dodecamer, which are large (Dickerson & Drew, 1981).

(2) The magnitudes of the local variations in structure are much less marked than in either the crystal structure of the dodecamer or the refined solution structure of the undecamer. Moreover, there appears to be no systematic difference between the conformation parameters of the purine and pyrimidine residues.

(3) All the base-pairs are propellor-twisted and this is most marked for the $G_2 \cdot C_{11}$ and $C_5 \cdot G_8$ base-pairs, which have average propellor twist angles of 13° and 12°, respectively. This large degree of propellor twisting is associated with a large negative base roll (i.e. opening towards the major groove) of −7° for the adjacent base steps A_4pC_5 (strand 1) and $A_{10}pC_{11}$ (strand 2).

(4) As in the crystal structure of the *B* DNA dodecamer, the extent of base overlap is variable. This can be seen from the stereo drawings of the five individual base-pair steps viewed down the helix axis (Fig. 4). Base overlap is most extensive for the two Pur_pPyr steps (base-pair steps 2 and 4).

In this letter we have shown that the solution structure of an oligonucleotide, namely the self-complementary *B* DNA hexamer 5′d(C-G-T-A-C-G)$_2$, can be successfully refined from two different *B* DNA structures on the basis of experimentally determined inter-proton distances to yield virtually identical refined structures. The results presented here clearly show that the way is now open to probe the molecular details of the three-dimensional structures of oligonucleotides in solution at a resolution comparable to that attainable by X-ray crystallography. The data base afforded by the refined structure of the *B* DNA hexamer as well as that of the *B* DNA undecamer (Clore & Gronenborn, 1985*b*) is too small to enable one to deduce any general rules as yet for the sequence dependence of local structural variations in solution. Such rules should emerge in the fullness of time when the solution structures of other oligonucleotides, refined on the basis of interproton distance data, become available.

G. Marius Clore, Angela M. Gronenborn

Max-Planck Institut für Biochemie
D-8033 Martinsried bei München, F.R.G.

David S. Moss and Ian J. Tickle

Department of Crystallography
Birkbeck College
Malet Street, London WC1E 7HX
U.K.

Received 8 October 1984, and in revised form 16 April 1985

References

Arnott, S. & Hukins, D. W. L. (1972). *Biochem. Biophys. Res. Commun.* **47**, 1504–1509.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *J. Comput. Chem.* **4**, 187–217.

Clore, G. M. & Gronenborn, A. M. (1984). *FEBS Letters*, **172**, 219–225.

Clore, G. M. & Gronenborn, A. M. (1985*a*). *J. Magn. Reson.* **61**, 158–164.

Clore, G. M. & Gronenborn, A. M. (1985*b*) *EMBO J.* **4**, 829–835.

Dickerson, R. E. & Drew, H. R. (1981). *J. Mol. Biol.* **149**, 761–785.

Gronenborn, A. M., Clore, G. M. & Kimber, B. J. (1984). *Biochem. J.* **221**, 723–736.

Haneef, I., Moss, D. S., Stanford, M. J. & Borkakoti, N. (1983). *RESTRAIN User Guide*, Birkbeck College, London.

Haneef, I., Moss, D. S., Stanford, M. J. & Borkakoti, N. (1985). *Acta Crystallogr. sect. A*, in the press.

Kuzmich, S., Marky, L. A. & Jones, R. A. (1982). *Nucl. Acids Res.* **10**, 6265–6271.

Levitt, M. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 640–644.

Rhodes, D. (1982). In *Topics in Nucleic Acid Structures* (Neidle, S., ed.), part 2, pp. 287–304, MacMillan Press, London.

Edited by M. Moody